DOCUMENT RESUME

ED 223 971 CS 006 881

AUTHOR Spangler, Katy
TITLE Readability: A Review and Analysis of the Research.
PUB DATE Jan 80
NOTE 51p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cloze Procedure; Comparative Analysis; Literature Reviews; *Readability Formulas; *Reading Research; Research Design; Research Methodology; *Test Interpretation; Test Reviews; *Test Validity

ABSTRACT

This paper reviews seven research studies on the subject of readability. The first study reviewed is itself an extensive review of 30 readability formulas described by George A. Klare. Of these, five formulas considered to be interesting, unusual, or classic on the basis of high validity, simplicity or complexity, common or uncommon variables, and other unusual features were chosen for analysis. The five formulas include those by I. I. Lorge (1939), G. D. Spache (1953, 1974), W. B. Elley (1969), J. R. Bormuth (1966, 1969), and Harris-Jacobson (1975). In addition, the initial research on cloze procedure by W. L. Taylor (1953) is reviewed to give balance to the overview of readability research. Each review consists of an analysis of the research backing the formulas, specifically the theoretical framework, the research design, the results, the author's evaluation, and a summary including comments on the usefulness and the face validity of the formulas. After the reviews, a synthesis of the studies attempts to answer the following questions: (1) What is readability and how is it calculated? (2) How good is the research on readability--what are its strengths and limitations? (3) How do readability measures compare? and (4) What are some practical implications for use of these formulas? (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED223971

READABILITY: A REVIEW AND ANALYSIS OF THE RESEARCH

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Katy Spangler

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Katy Spangler

University of Washington

January, 1980

CS006881

INTRODUCTION

This paper consists of reviews of seven research studies on the subject of readability, plus a synthesis of these studies. The first study is an extensive review of 30 readability formulas by George A. Klare (1974-75). In order to get a comprehensive picture of all the material he covered, I charted information on the 30 formulas. (See Chart 1.) The chart focuses on the variables of each readability formula, validation, the criterion used for correlation, and the levels/gradations that are used to report the results. From this chart I picked and reviewed five formulas which I consider either interesting, unusual or classic, on the basis of high validity, simplicity or complexity, common or uncommon variables, and other unusual features. The formulas reviewed are:

1.) Lorge (1939) -- a classic

2.) Spache (1953, 1974) -- another classic, well respected

3.) Elley (1969) -- high validity, unusual variable

4.) Bormuth (1966, 1969) -- complexity and completeness

5.) Harris-Jacobson (1975) -- recent, high validity, "in" formula

In addition, the initial research on cloze procedure (Taylor, 1953) is reviewed in order to give a balance to my overview of readability research because it illustrates a measurement rather than a predictive technique. Each research report consists of an analysis of the research backing the formulas, specifically, the theoretical framework, the research design, the results, my evaluation, and a summary including comments on the usefulness and the face validity of the formulas.

After the research reports, a synthesis will attempt to answer these questions:

1.) What is readability and how is it calculated?

2.) How good is the research on readability--what are the strengths and limitations?

3.) How do readability measures compare?

4.) What are some practical implications for use of these formulas?

Note that the Klare, Standal (1978), Britton and Lumpkin (1977) and Pescosolido and Gervase (1971) papers served as resource lists and summaries, but are not the bases for the judgements made in this paper.

Klare, George A., "Assessing Readability." Reading Research Quarterly Fall, 1974, 10:1, p. 62-102.

Chart 1

Summary of Klare Article

Name & Date	Description	Variables	Validity	Criteria	Levels	Comments
Lorge, 1948	Regression Eq.	-Sentence length -Dale hard words -Prep. phrases	.77	McCall-Crabbs 50,75% comp.	3-12 gr	uses 739 Dale list
Flesch, 1948 "Reading Ease"	Regression Eq.	-Syl/100 words -words/sentence	.70	McCall-Crabbs 50,75% comp.	Adult	
Dale-Chall 1948	Regression Eq.	-% words on Dale list -av. sent. length	.70 .71	McCall-Crabbs-1925 50% comp McCall-Crabbs-1950	Adult	uses 3000 word list Recalculated 1958 Has other versions
Farr-Jenkins-Paterson, 1951	Regression Eq.	-# 1-syl words/ 100 -av. sent. length	.93-.95 .58	Flesch Rdg Ease McCall-Crabbs-1950	Adult?	uses "Reading Ease Index"
Fog (Gunning) 1952	Regression Eq.	-Av. sent. length -%words over 3 syl	? .59	Flesch Rdg Ease McCall-Crabbs-1950	Adult?	
Spache, 1953	Regression Eq.	-Av sent length -# words not on Dale list	.818 .95	Level classroom use of books (w/ revised word list)	Gr 1-3	uses Dale 769 list
Devereaux (Smith) 1961	Regression Eq.	-characters/word -sent. length in words	--	basal readers	Gr 4-12	Validity rises with grade level
Jacobson, 1961, 1965	Regression Eqs.	-words/clause -math terms -word freq -scientific terms -word length	.67-.70	Test scores on text passages	High Sch-College	For Chemistry & Physics

Name & Date	Description	Variables	Validity	Criteria	Levels	Comments
Botel, 1962		-Word difficulty		Vocab in basic texts, basals & magazines	PreP-12	uses "graded vocab list"
Rogers, 1962 "Listening Comp"	Regression Eq.	-Av idea unit length -words not on Dale 3000	.727 (SEM 2.37)	Grade level of sample	K-12	
Danielson-Bryan 1963	Regression Eq.	-characters/wd -characters/sent	.575	McCall-Crabbs	?	Designed for automation
Fry, 1965	Graph	-syl/100 words -wds/sentence	?	"primary & sec. materials"	K-12	
Coleman, 1965	Regression Eq.	-1 syl wds/100 wds -sent/100 wds -pronouns/100 wds -prep/100 wds	.86 .89 .90 .91	Cloze tests	General	4 formulas use unclear
Fang, 1966-67 "Easy Listening"		-# of syl above 1/wd in a sent	.96	Flesch R.E., TV scripts & newspaper	"Mass Listenability"	No gradations made
Shaw, 1967	Regression Eq.	-science & math wds -%frames that were response frames -av # review frames	--		4th grade	Programmed bases on indiv. errors / subject
Readscore (undated)	Regression Eq.	-av sent length -av SD wd length -av # preps. -"wd length factor" -av # connectives	.72	McCall-Crabbs	?	For computor
Elley, 1969	Correlation??	-Noun frequency	.90	Judges' criteria	?	Uses regional word list & conversion table

Name & Date	Description	Variables	Validity	Criteria	Levels	Comments
McLaughlin 1969 "SMOG"	Simple reg eq. uses square root	-word length -sent. length (# long words)	SEM=1.5	McCall-Crabbs 100% comp	?older	High gr levels - can be corrected
Bormuth, 1969	24 Reg. eqs.	-tries up to 169 diff. variables	.52-.83	Cloze tests 35,45,55% comp	1-College	Experimental-2600 subjects Simple to complex
Aquino	7 Reg. eqs.	-Latin based syls -complex N phrases -"although" clauses -preposed adjs -Latin suffixes	.90+	Cloze tests-College students	?older	Linguistic variables Imps. for rewriting
Mugford, 1970	Chart/graph	-word length -sent. length -repetition	--	Ratings of authors, teachers 7 publishers	k-adult	
Jablonski, 1970 "Boeing"	Profile	-word complexity -personal ref. wds. -word length -sent length -verbal connectives -modifiers	"passing"	"Appropriate rdg levels for 4 levels"	--	4 levels, used for rewriting
Henshall, 1971	Reg eq. composed of 2 other eqs.	-Farr-Jenkins-Patterson -Gunning FOG	?	??	College	for shorthand
Botel & Granowsky 1972 "Syntactic Complexity"	?	-syntactic complexity	--	--	--	in process
Caylor, Stict, Fox Ford, 1973 FORCAST	Regression eq.	-one syl words	??	Cloze tests 35%	older?	For use in Army
Lensear Write	Regression eq.	-one syl. words -sent. length	no	no	Average adult	No gradations

Name & Date	Description	Variables	Validity	Criteria	Levels	Comments
Morris, 1972 Clear River	Chart	-words/sent -wds/punc pause	no	no	Average adult	
Dameŕst Clear Index, 1972	Chart-index	-one syl words -sent length	no	no	?	Scale 0-100
RIDE	Scale	-av # letters/ word	less than controls	Rdg Ease, Dale-Chall, A.R.I.. SMOG, Fry, cloze	K-College	5 levels -- too general
Harris-Jacobson 1973 "Primary Reading Formulas"	Regression eqs.	??	.90 SEM=.38, and .71	Graded basals	Gr. 1-6	

Lorge, Irving I. "Predicting reading difficulty of selections for children." The Elementary English Review October, 1939, 16, 229-233.

Lorge, Irving I. "The Lorge and Flesch readability formulae: a correction." School and Society February 21, 1948, 67, 141-142.

Lorge, Irving I. The Lorge Formula. New York: Teachers College Press, 1969.

Theory

Lorge theorized that readability was reportable on a criterion of difficulty as a function of the internal structure of reading passages in his 1939 article. His research was the first to develop a readability formula applicable to children's reading passages. It was based on an earlier (1935) formula for adult passages by Gray and Leary. The research question in Lorge's study was: which factors of internal structure affect the difficulty of children's reading passages?

Design

The research consisted of calculating the correlations between grade level scores of children's reading passages and 6 factors of internal structure of these passages. The materials used for study were the "McCall-Crabbs Standard Test Lessons in Reading" (1925). Three grade level scores from the 378 passages were used for correlations: normed grade level scores for 25% comprehension (C_{25}), for 50% (C_{50}) and for 75% (C_{75}). Multiple correlations were made with 6 factors of internal structure. Based on the Gray-Leary paradigm, these were:

1.) number of personal pronouns/100 words
2.) number of words not on Dale 769 List
3.) average sentence length in words
4.) percent of non-Dale 769 words to total words (this seems like the same as #2 to me.)
5.) number of prepositional phrases/100 words
6.) weighted index of words for frequency in Thorndike (Word Frequency Book?)

Results

The results of correlations with the C_{50} scores and all 6 variables was r=.7722. Other correlation coefficients reported for C_{50} with combinations of less than 6 variables ranged from .6949 to .7721. A final simplified formula using only three variables and with r= .670 was reported in 1948:

GL = .10X + .06Y + .10Z + 1.99

where GL = Grade Level
X = number of words not on Dale 769 list
Y = average sentence length in words
Z = number of prepositional phrases/100 words

According to Lorge's tables, the validity of this formula was the highest of its kind reported as of 1948. The most significant factor of reading difficulty was found to be vocabulary (as is represented in the word list variable). The other significant factor was termed "structural elements" (represented by sentence length and density of prepositional phrase variables).

Evaluation

On the surface, the design of this research seems fine. However, the design was based on the premise that the grade levels reported in the McCall-Crabbs passages were representative of actual reading comprehensibility of the passages by children. I decided to examine these materials to see what they were like.

The Lorge reserach was based on the 1925 McCall-Crabbs passages. I was only able to examine a third edition (1961), Level E, Grades 7-12. No norming, reliability or validity information was available. The test booklet consisted of 78 paragraphs of 160-22 words followed by 9-12 literal questions. Directions were to read the material and answer the multiple choice questions within

3 minutes. This material was normed on subjects grades 2-12: average number of questions correct by grade level is reported for each paragraph. In the Lorge research, a passage was graded at the level where 50% of the questions were answered correctly. For example, Passage 49, on which the average child in Grade 6, 8th month got 5/10 questions right is graded 6.8. Thus, the difficulty or grade level is 6.8, according to Lorge. This criterion of "difficulty" or comprehensibility seems nacurate to me. First, the paragraphs seemed to be normed on students' timed literal responses and organizational skills. Thus, I think the paragraphs may represent "too easy" a level of comprehension. Second, the criterion for a grade level was only ½ "comprehension. This makes the passages seem even "easier". Third, all the material consisted of single paragraphs which were too short and too estranged from the "real world" to constutute measures of comprehension. Thus, while I feel the research design is fine, the material on which it is based was inappropriate. Hence, the grade levels may be erroneously assigned. (N.B. Lorge does supply a correction factor for C_{75} if desired. This heightens the validity of his research, but still does not erase the problem.) Since the Lorge formula is a "granddaddy" of may readability formulas, the idea of its validity will be discussed further in the synthesis portion of this paper.

Summary

The Lorge formula combines the factors of vocabulary difficulty and sentence complexity. These are the two traditional factors in readability formulas (Standal, 1978), however they are measured by 3 variables, where preposition frequency and sn-tnece length are both measures of sentence complexity. Thus, the Lorge formula is a 3-variable regression formula which uses a word list. According to Klare, it was "the first modern, easy-to-use formula". It has been revised in 2 ways: 1.) by using word length instead of the Dale list for the vocabulary factor, and 2.) by dropping the prepositional phrase factor. It has been recalculated on the 1961

McCall-Crabbs paragraphs.

Sample

The calculation of the Lorge formula was done on the short story "Dreams Can Come True" by Jesus Cortez. The formula is time-consuming and cumbersome, because both prepositions and hard words must be counted. I did the formula in two forms; that of the 1948 correction, and that of the 1969 handbook. The 1948 formula is poorly worded and should be reworded as follows:

$$GL = .10X + .06Y + .10Z + 1.99$$

where GL = Grade Level

X = Ratio of non-Dale 769 words to total words in passage

Y = Average sentence length in words

Z = Ratio of prepositions to total words in passage

This calculation resulted in a 3.9 Grade Level. A second calculation was done using the 1969 handbook, which resulted in a 3.8 Grade Level on the same material. The handbook gives specific rules for word counting, the Dale 769 Word List, and and worksheet format to simplify calculations. This is easier to use for the non-mathematically inclined.

Spache, George D. "A New Readability Formula for Primary Grade Reading Materials" Elementary School Journal, March 1953, 53, p. 410-413.

Spache, George D. Good Reading for Poor Readers Champaign, Illinois, 1974.

Theory

Spache (1953) wanted to design a readability formula that could be used for primary level materials. Agreeing with Lorge, Flesch and Dale & Chall, Spache theorized that readability is affected by vocabulary and sentence length. While acknowlegding their validity, Spache chose not to employ number of affixes or extended word lists in his formula due to the simplicity of primary level materials.

Design

In this study, a simple multiple correlation was made between 2 measures of passage difficulty and grade level:

1.) Average sentence length

and 2.) Number of "hard words" not on Dale 769 list

with 3.) grade levels of textbooks

The materials were 224 samples of 100 words each from 152 "textbooks in common use" in grades 1-3. Determination of grade level of the books was done simply by arbitrarily assigning a number to each book on the basis of the grade level at which it was used. Spache reports the following gradations:

1.2	Preprimer
1.5	Primer
1.8	First reader
1.9	First reader, second level
2.1	Second reader
2.4	Second reader, second level
2.7	Second reader, third level
3.3	Third reader
3.7	Third reader, second level

Grade levels were also assigned to 23 social studies, science and health texts; the remainder of the texts were basal readers.

Results

The correlation coefficient of the Spache study was .818. The finalized regression formaula (1953) was:

Grade Level of Textbook = .141X + .086Y + .839

where: X = average sentence length

Y = percentage of words not on Dale 769 list

Spache revised his formula in 1974. The basic design was exactly the same, but the following modifications were made:

1.) The word list was revised and expanded based on the changing trends in expanded basal reader vocabulary and research on primary-age children's vocabulary.

2.) 100 new samples from supplimentary books, basal readers, science and social studies texts were analyzed.

The resulting formula:

Grade Level = 1.21X + .082Y + .659

has a multiple correlation coefficient of .952, S.E.M. = .2. This was and is the highest validity ever reported in readability research.

Evaluation

Spache's research was natable in that his material was from real classrooms; what real children were reading. This choice of material also has its drawbacks: in the possible imprecision of assignment of grade level (apparently done by personal judgment of 2 graduate students after classroom observation), and the possible unrepresentativeness of the children and the books used. (Neither the location of the schools, the number of classrooms observed nor the books used are reported.) The original formula had the drawback of using only textboods for evaluation--in the revised form, children's trade books (I assume this is what "supplimentary texts" means) were included.

The expanded wordlist in Spache's 1974 revision has 1,034 words. While longer, it appears to be more representative of primary vocabulary and has

a research base.

Spache (1974) notes that his formula has been validated by field research. The Spache formula was compared to pupil performance on IRIs with a favorable .70 degree of correlation. It was found that only 3 100-word samples are needed to use the Spache formula accurately.

The Spache formula may be most accurate in evaluating text books and least useful for trade books. Both the word list and the materials analyzed appeared to be predominately texts. One might expect trade books to have larger vocabularies, higher interest levels and possible better picture clues by grade level than textbooks. This might change readability ratings.

A drawback in Spache's research is imprecision of reporting. While both articles cited are merely summaries that are to be used practically, a detailed technical report including the materials sampled, the criteria for grade level assignment, the types of classrooms visited, etc., would add further credence to Spache's research.

Summary

Spache has devised what is probably the most popular readability formula for primary reading materials. It combines vocabulary frequency (using a word list) and sentence length in a 2 variable regression formula with extremely high validity. It's up-to-date revision takes into account both hard research and (hopefully) modern materials. The formula is limited to use with primary materials.

Sample

The calculation of the Spache formula was done on an unpublished short story "Dreams Can Come True" by Jesus Cortez. I did the calculation in two forms: first with three 100 word samples, receiving a readability of 2.4. Second, I did the calculation for the entire story, rating 2.0

The calculation is time consuming since all words have to be checked on the word. Since the formula is not made for analyzing all the words in a long passage, I think the 2.4 rating is the more accurate of the two. Spache's Revised Word List seems representative of primary vocabulary.

Elley, Warwick B. "The Assessment of Readability by Noun Frequency Counts" Reading Research Quarterly, 4, Spring, 1969, p. 411-427

Theory

Elley's unique method of determining readability is based on a combination of three theories. First, meaningfulness of vocabulary is directly related to word frequency: ie., the more common the word, the easier it is to read and comprehend. This theory is supported by traditional readability and verbal learning research. Second, nouns, as a class, are the most "weighty" part of speech--the most related to meaning. This is because a.) they compromise the largest proportion of children's speech and adult writing; b.) they are the least redundant part of speech (least often correctly answered on cloze tests, and least self-corrected by children reading out loud), and c.) they tend to be of the rare, obscure or abstract type in "difficult" prose. Third, teachers' ratings of readability of passages appears to be the most reliable criterion available. While teacher ratings and clozed tests correlate highly ($r = .91$), Elley found that teacher rating had consistently superior correlations with various other measures of readability. Thus, Elley's hypothesis for measurement of readability: that frequency of gradiated "hard" nouns in children's reading materials will correlate with teacher rating of passage difficulty.

Design

Elley's research consisted of a multiple correlation between 12 indices of readability and average readability ratings of passages. The indices of readability were:

Noun frequency level
Noun and adjective frequency level
Noun and verb frequency level
Lorge formula

Syllables per 100 words
Dale-Chall formula
Proportion of abstract nouns
Proportion of unfamiliar words
Verb frequency level
Adjective frequency level
Sentence length
Ratio of prepositional phrases

In these measures, "frequency level" was determined by using the New Zealand Council on Educational Research Alphabetical Spelling List. This graded list of 2700 words was propared on the basis of the words most frequently appearing in children's writing. In other words, Level I would list the most common words, Level II, less common, and so on, up to Level VIII. "Frequency level" was determined by the average level of the words appearing in a passage divided by the total number number of words of that class in the passage. For example, if noun frequency is to be determined, for all the nouns appearing in a passage, the level from the word list for each of those nouns is determined and then averaged, then divided by the total number of nouns in the passage. The resulting "noun frequency level" is the indice of difficulty.

Readability ratings were given by teachers (on the basis of eye-balling the materials) on 5 series of 150 word passages, a total of 58 passages in all. In the last three series, "top-stream" 12-15 year old student ratings, which correlated highly with teacher ratings, were included. The number of raters per series varied from 10 to 50. As mentioned above, teacher ratings, while correlating highly with cloze measures, were consistently the most reliable.

Results

Noun frequency level and average judge ratings had the highest average correlation, $r = .90$. Mean correlations for the other measures (in the rank listed in the design section) ranged from .89 for noun and adjective frequency

to .41 for ratio of prepositional phrases. All correlations were quite high: for the Lorge formula, r = .87, for the Dale-Chall, r = .85.

The resulting formulas to estimate readability is then as follows:

$$AFL = \frac{X}{Y}$$

where: AFL = Average Frequency Level

X = total of levels from NZCER list for all nouns in the passage

Y = total number of nouns in the passage (must be over 20)

Evaluation

The advantages of Elley's formula are many. First, its reliability is high. Second, it is extremely simple to calculate, since it has only one variable. Third, if graded word lists are available, it can be used in various English-speaking nations with specific consideration of the dialects and special regional usages of those locales. (Elley used word lists from New Zealand, where the experiment was conducted, from England, and from New York.) Fourth, the entire system is set up based on valid sources of information: the word lists are based on children's writing, and the difficulty levels are based on estimates considered better than cloze measures. Fifth, the research backing Elley's formula is careful and complete: the resulting formula is original and unique.

There are two glaring practical disadvantages, unfortunately. The first is that a proper word list must be available. It must be grade-leveled, which no doubt is done on some subjective basis. It must be based on actual children's writing, and it should have more than 2000 words. It should be regional in order for it to be most accurate. If I use the New York list, I may be computing readability for New York City children, but not for children in Seattle. (This is a problem, however, of all reading materials, and thus all readability formulas.)

The second problem is that the resulting ranking of passages are relative ranking only, and are not grade levels. Elley includes a grade level conversion table, based on publisher labeling of basal books and teacher gradings. Right here, we can see all sorts of subjectivity being introduced, lessening credibility of the .90 reliability coefficient. The table gives age levels as final ratings, which may be appropriate for New Zealand, but less so for the United States.

So, here is the unfortunate weakness of this otherwise outstanding study: of what practical value is the fact that the noun frequency level measure accurately ranks materials judged consistently by teachers? What teachers (and writers) need (aside from the confidence that they are good judges of readability) is actual grade level equivilents of the materials they take the time to analyze.

Thus, the Elley formula, while both reliable and valid, can only, so far, be used to rank the difficulty of children's passages. Further analysis of the formula could make it more useful: it seems to me that by correlating the noun frequency level directly with materials chosen by teachers or children as representative of the various grade levels, a useable, simple, and even dialectically sensitive guideline could be formulated.

Summary

The Elley formula ranks children's reading passage difficulty by level frequency of hard nouns per total nouns. A graded word list of at least 2000 words is necessary. Results of this study apply to New Zealand: further analysis would be necessary to validate the process for readability in the United States.

Sample

I was unable to use the Elley readability process due to inavailibility of a regional vocabulary list. This would be the major drawback of his formula.

Bormuth, John R. "Readability: A New Approach" Reading Research Quarterly, 1:3, Spring, 1966, p 99-132

Bormuth, John R. Development of Readability Analyses Final Report, Project #7-0052, Contract # OEC-3-7-070052-0326, USOE Bureau of Research, HEW. March, 1969, ERIC Document # ED 029-166.

Theory

The basic theory backing Bormuth's (1966) massive research project on readability was that the cloze procedure gives an accurate and direct measure of readability. Therefore, a number of linguistic variables were correlated to cloze difficulty to test their predictability of word, independent clause, sentence and passage difficulty. Several technical advances made this research possible. First, the cloze test is versatile, direct and accurate for use as a grade level comparison device. Second, new linguistic variables capable of careful analysis have been developed. Third, computor technology makes analysis of many variables and a wide range of materials possible. Thus, Bormuth posed 5 theoretical questions:

1.) Are linguistic features linearly related to readability? (Traditional readability formulas indicate that they are.)

2.) Do the factors affecting readability change with the reading ability of the reader? (Are different formulas necessary for beginning and fluent readers?)

3.) Can readability be computed for language units as small as the word, sentence or independent clause? (This would be useful for analysis of test questions, headlines, directions, etc.)

4.) Do modern testing and linguistic analysis techniques imporove readability validity? (Bormuth states that traditional readability formulas are "crude" in terms of today's technoloty.)

5.) Which new linguistic variables (word depth, letter redundancy, independent clause frequency, letter counts, part-of-speech analysis using structural grammar) affect readability?

Design

Materials were twenty unfamiliar 275-300 word passages, all by different authors, and evenly distributed in content areas and difficulty. Cloze tests

deleting every fifth word were made so that, with 5 forms, all words were subject for analysis. (First and last sentences of the passages were not analyzed.) All total, 5,181 words, 405 independent clauses, and 365 sentences were tested.

The subjects were 4th-8th graders from Southern California. They were considered academically, socio-economically and racially representative of the United States. Their overall mean on the California Reading Test was 5.5, ranging from 1.4-12.7. The children were divided into 5 matched form groups (n = 135 for each). Each form groups was divided into 5 significantly different ability level groups.

Each child completed all 20 untimed cloze tests from one form over a period of 10 school days. The tests were systematically hand scored, allowing only for spelling errors.

The dependent variable was the prooprtion of correct responses on each deleted word. This dependent variable was calculated for 4 language units: 1.) word; 2.) independent clause (mean word difficulty per independent clause); 3.) sentence difficulty (mean word difficulty per sentence); and 4.) passage (mean word difficulty per passage).

These were correlated to the following indices of reading difficulty:

1.) Word Depth (a measure of within-word and within-sentence information)
2.) Letter Redundancy (a measure of word and syntactic difficulty)
3.) Independent Clause Frequency (based on children's language patterns)
4.) Letter Counts/Unit
5.) Parts of Speech/Unit (by traditional and structural systems)
6.) Word Frequency (based on Lorge-Thorndike word list
7.) Syllable Count/Word
8.) Sentence Length
9.) Proportion of Words Not on Dale 769 List

10.) Proportion of Words Not on Dale 3000 List

11.) Ratio of Prepositional Phrases to Total Words

12.) Proportion of Total Words Contained in Prepositional Phrases

13.) Independent Clause Length

Analyses and Results

Question 1: Are linguistic features linearly related to readability? A linear and a non-linear regression equation was calculated for each variable, than they variance accounted for by each was tested by the F test of linearity. F ratios were significant for 4 variables: syllables/word, letters/word, word frequency and Dale 3000. This means that relationships between these factors (and possible others) and reading difficulty are not linear. For example, in letters/word, as the words get longer, they do not necessarily get progressively harder at the same rate. The Dale 3000 test results confirm Chall's assertion that this formula is accurate only for a limited range of materials. As a result, quadratic equations might be more adequate for precise readability formulas.

Question 2: Do the factors affecting readability change with the ability of the reader? Reading levels of the subjects had been employed to make 5 different ability groups. Likewise, for each readability variable, 3 levels of difficulty were defined. A 3x5 analysis of variance, while significant for all main effects, revealed only one significant interaction. This means that readability formulas would be equally valid to use for differing ability readers, including the more able readers.

Question 3: Can readability be computed for small language units? Stepwise multiple correlations were computed at the word and independent clause levels. Variables were systematically added into the regression

equations until adequate r values were obtained. At the word level, r = .505 with 4 variables. At the independent clause level, r = .665 with 6 variables. Bormuth concluded that the moderate r values indicated that by adding more variables and making the regression equations curvilinear, accurate readability assessments of small language units could be made by computor.

Question 4: Do modern testing and linguistic analyses improve readability validity? Again, stepwise multiple correlations were comptuted, and then entered into regression equations. For a 4 variable equation including both traditional and new factors, an r value of .934 was obtained. Bormuth concluded that by adding more variables, more precise measures of passage difficulty could be obtained and that .99 correlations are not impossible in the future.

Question 5: Which new linguistic variables affect readability? The power of the 5 types of new variables was examined by reviewing their correlations to cloze difficulty and to other variables, and by analysis of variance. Word depth, letter redundancy, letter counts, and part-of-speech ratios (especially pronouns/conjunctions where r = .81) all appeared to be useful variables. Bormuth stated that sentence length and complexity need not be synonomous, and can be measured separately. It was also indicated that different variables are more applicable to different units of analysis.

The results of the 1966 study were used to direct Bormuth's next research project, which was to provide readability formulas based on cloze difficulty and his findings described above. Bormuth conducted another massive data collection similar to the first, using readers ranging from grade level 3 to graduate students. Data were analyzed on 50 linguistic variables, at 4 unit levels, for mean, 35%, 45% and 55% cloze scores.

Cloze level scores were correlated to 75-90% comprehension question scores on the passages to determine grade level equivilents. Many formulas resulted, some having up to 22 variables. While not really representative, I chose what seemed the most practical and easy-to-use formula for passages:

$$C_{45} = 1.85 + 1.31X + .05Y - .0001Y^2$$

where C_{45} = Grade placement with 45% cloze score

X = Ratio of letters to words

Y = Ratio of letters to minimum punctuation units

This particular formula, for manual use, has $r = .84$.

Evaluation

First, note that the above equation takes into account some of Bormuth's (1966) original findings. It is a quadratic equation, it is not restricted to any ability or difficulty level, and it analyzes two old familiar variables (vocabulary and syntactic complexity) in a new manner. Also note that the validity coefficient is high, and the resulting score is a direct grade level measure of student comprehension. (Grade levels are represented by equations for 35%, 45% and 55% cloze comprehension. I chose the middle as the most representative measure.) So, when a passage is analyzed using this formula, the resulting score should indicate at what grade level proficiency a reader functions in order to have good comprehension of that passage. Thus, the formula apparently has high validity to measure reading comprehension.

While the simplest formula is discussed here, even it has its drawbacks. It is time-consuming to compute: machine analysis would be desireable for the letter counting. On the other hand, word lists are not needed. See below for a discussion of the formula's comparibility with other formulas in terms of difficulty.

The Bormuth formulas are obviously based on extensive and careful re-

search. Bormuth's organization and interpretation of such a vast amount of data is especially commendable, as is his application of the theoretical to the practical. His knowledge and use of modern linguistic theory is creative and original. Also, development of formulas for varying units of language analysis is a creative innovation. The fact that Bormuth based his research on actual reader performance by using cloze tests makes his research seem more credible than most readability research.

Bormuth's two studies point out the extreme complexity of the prediction of difficulty of written language. The fact that so many variables are significant should lead to accurate readability assessment more as a computor problem than a teacher's job.

Summary

The Bormuth research provides numerous simple to complex formulas for the assessment of word, independent clause, sentence and passage difficulty, for all reading levels. The regression equations are curvilinear and have 2-22 variables. The formulas are not generally appropriate for manual calculation.

Sample

The short story "Dreams Do Come True" by Jesus Cortez was analyzed using a simple manual formula developed by Bormuth. The formula,

$$C_{45} = 1.85 + 1.31\left(\frac{\text{letters}}{\text{words}}\right) + .05\left(\frac{\text{letters}}{\text{MPU}}\right) - .0001\left(\frac{\text{letters}}{\text{MPU}}\right)^2$$

yielded an average grade level of 8.6. The score is obviously very high, but was reconfirmed by several more calculations. I can only conclude that 1.) this formula is extremely tedious to calculate and 2.) perhaps due to its unusual validation technique, rates passages much higher than other readability formulas.

Harris, Albert J., and Edward R. Sipay How to Increase Reading Ability, 6th ed. New York: David McKay Co, Inc., 1975, p. 658-675.

Theory

This formula is based on a theoretical background provided by numerous researchers: that passage difficulty is based on a combination of sentence length measures and word difficulty measures. Thus, these two variables, when weighted properly, provide the best measure of readability. Furthermore, the Harris-Jacobxon formulas are based on two other theoretical assumptions: 1.) that current children's basal reader series provide a good match between reader and passage difficulty, and 2.) that factors of readability vary with age/grade level/ability of the reader.

Design

Six basal reader series were used as the sample. Each series had ten levels (with three pre-primers counted as one level). From each of the six series, ten smaples of approximately 200 words were taken from each level: 6 x 10 = 60 smaples/level x 10 levels = over 600 smaples totaling about 135,000 words. These samples were scored by computor for seven variables which are scoreable either by hand or computor, and entered into a multiple correlation program. The computor chose the best two and three variable combinations for four different grade level combinations.

Results

The two "best" formulas are as follows:

Formula 1, grades 1-3:

Predicted Score = .094X + .168Y + .502

Formula 2, grades 4-6:

Predicted Score = .140X + .153Y + .560

where X = % words not on Harris-Jacobson Short Readability List

Y = Average sentence length

Predicted Score = number corresponding to level of reader (Primer High Third, Sixth, etc) on chart provided.

The validity for Formula 1 is R = .90, SEM = .384; for Formula 2, R = .90, SEM = .714. This is an extremely high correlation in terms of readability formulas, and low standard errors of measurement. Although the formula for upper level books has a higher SEM, about double that of the first formula, this is compensated for because the lower level formula has 2 levels/grade, while the upper level formula has one level/grade. Thus, both formulas have a standard error of about .7 of a grade level from the "true" readability level. In other words, most of the time, you can assume that the measured grade level is accurate within ±.7 months of the actual grade level.

Evaluation

The Harris-Jacobson formulas reflect a modern approach to readability. They were determined by computor, are highly valid and reliable. They reflect an approach similar to Bormuth's in that a number of variables and a large sample were considered in order to determine an accurate hand-scoreable formula.

There are also some interesting disagreements with the Bormuth research. First, the formulas are not quadratic, as Bormuth predicts. Second, the formulas are different for grade levels, which Bormuth does not support.

A thire disagreement with the Bormuth research, and a general question of the validity of this (and many other readabilty studies) is that the criterion for grade level used by Harris and Jacobson is not based on actual children's performance, but instead on arbitrary labeling of grade level on current basal readers. If "current" basals are considered "hard", which I know some are (for example, Houghton-Mifflin) then these scores may not reflect the way children read, but instead the way publishing companies think

children read. Thus a score with the Harris-Jacobson formula correlates with books which may or may not correlate with average reading ability. A more direct correlation to a criterion such as measured ability of pupils to comprehend passages may be desireable.

Rumor has it that the Harris-Jacobson formulas are the latest of "what's in". This may be true of what's "in" for publishers, but seems less useful and accurate for teachers. Another drawback from the researcher's point of view is that the published description of this new formula does not give details: for example, the six basal series used, the ten original variables tested, and the resulting correlation matrices are not included in the article. In all fairness, the article reviewed appears in a textbook, and long-winded theoretical description is inappropriate, but a journal article with details would be useful.

Summary

The Harris-Jacobson formulas consist of two 2-variable regression equations, one equation for primary grades, the other for grades 4-6. The equations are based on a word list for vocabulary difficulty and sentence length for syntactic complexity. The formulas, while employing traditional variables, are more valid and reliable than the older formulas, perhaps due to computor technology. They are hand scoreable.

Sample

Following instructions, I calculated readability from three 200-word samples in the short story "Dreams Do Come True" by Jesus Cortez. Since thestory seemed at about 3-4 grade level, I tried both formulas, arriving at 2.9 (low third) with Formula 1, and 3.4 (low third) with Formula 2.

The problem with this formula is the same as with all word list formulas -- it takes an extremely long time and is extremely boring to do the word counts. An advantage is that the results for both formulas were somewhat consistent.

Taylor, Wilson L. "Cloze procedure: a new tool for measuring readability." Journalism Quarterly 30. Fall, 1953, p. 415-433.

Theory and Definition

The cloze procedure is based on a theory of gestalt psychology that the human has a tendancy to mentally complete an unfinished mental pattern. In language, this means that a reader or listener will attempt to form closure on a message that has been multilated by deletions of letters or words. The cloze procedure per se is limited to written materials with systematic word deletions, and takes into account several more theories of language. One, language behavior depends on "total context". Two, language includes a great deal of redundancy. Three, each human develops unique language habits based on personal experience and circumstatce. Four, randomness insures proportionate representation of all possibilities of language/meaning closure. The cloze procedure is not a readability formula: it is a method to measure the similarity of author and reader's language patterns, implying that the more alike they are, the "easier" the material for the reader.

The Study and Results

Two experiments are reported in this article. In both, cloze tests consisting of typewritten sheets with 10-space word blanks for every nth deleted word were constructed. Subjects were university journalism students.

Experiment I Three passages were ranked using the Flesch and Dale-Chall readability formulas. Cloze tests deleting every 5th, every 10th and random words were constructed and administered in untimed sessions to 24 subjects to test the following hypotheses:

1.) Cloze tests would rank the passages the same as the Flesch and

Dale-Chall formulas. When formula rank and cloze scores were compared, this hypothesis was consistently supported.

2.) The relative cloze scores on the three passages would not differ with three different deletion systems (5th, 10th, random deletions). This hypothesis was confirmed by analysis of variance yielding significant f scores for differences between passages in all three conditions.

3.) The relative cloze scores on the three passages would not differ with different a.) deletion systems, b.) presentations orders, or c.) scoring methods. a.) The 10th word deltion system appeared to be less discriminating than the 5th word or random systems. b.) The order of presentation of passages make no significant differences. c.) The use of two scoring systems (exact word and synonym acceptance) made no significant differences.

Experiment II Five more passages were added to the three used in Experiment I. Cloze tests with the 7th word deletions were constructed. Flesch and Dale-Chall readabilities were calculated. Some of the new passages were deliberately chosen to "foil" the formulas: an "unintelligble" passage by Gertrude Stein, "characterized by short sentences and short familiar words" etc. Each cloze thes passage was completed by 18 subjects to test the following hypotheses:

1.) The results of Experiment I would be validated. This hypothesis was supported, even though the deletion systems were different.

2.) The rank order of 8 passages as determined by a pilot test with six subjects would be replicated by a larger sample. The hypothesis was supported, by a correlation of the rank differences, where r = .98, p .001.

3.) Cloze procedure would "handle" materials that traditional formulas could not. The rank arders for cloze, and the two formulas were different. The cloze procudure, for example, rated the Stein passage as hardest, where Flesch rated "very easy" and Dale-Chall rated 4-5th grade.

4.) Differences between cloze scores on the eight passages would be significant. Analysis of variance with p .001 confirmed this hypothesis.

In addition, the data were examined for two other considerations. One, to see if adjacently ranked passages were significantly different, t tests were computed. Results were inconclusive. Two, to test internal consistency, the passages were divided into fifths: each passage fraction was compared with the others for the high, middle and low scoring subjects. Internal consistency coefficients computed ranged from .56 to .95. It was concluded that shorter passages or fewer deltions would be equally accurate for all ability levels of subjects.

Evaluation

These experiments consisted basically of statistical analysis of a small number of cloze tests on a limited number of subjects. In spite of the "smallness" of the studies, they are significant in that Taylor's work initiated a tremendous amount of research and practical application of the cloze procedure. As far as I know, his findings have always been supported.

The cloze procedure, as discussed in this article, proved to be reliable and valid, simple to use, and more flexible than traditional reading formulas. Subsequent research and use has confirmed this.

As far as practical application, Taylor's work only goes as far as determining rank order of passage readability. Further research was necessary to correlate cloze difficulty of passages to reading ability.

Summary

The cloze procedure was introduced by Taylor in 1953. It is not a predictor of passage difficulty as are formulas. Instead, it is a measure of match between reader and author language. This introductory study dealt only with rank order correlations of cloze scores to passage difficulty as determined by traditional readability formulas and common sense. The study also detailed the construction of cloze tests which remains vir-

tually unchanged today, a quarter century later. The cloze procedure is widely used today in research on reading comprehension, as validation for new readability formulas, and in the classroom to match children and books.

SYNTHESIS AND DISCUSSION

What is Readability and How is It Calculated?

A primary concern of reading teachers and of people who read is that the consumer's reading proficiency and "knowledge of the world" match the difficulty of the material to be read. Assuming that the reader's skills and knowledge increase throughout the school years, it has become conventional to assign grade or difficulty levels to textooks and school library books. Readability is a prediction of the ease or difficulty of written language in relation to the abilities of the intended readers.

Researchers have developed readability formulas which can be systematically utilized to determine difficulty of prose passages. The actual process used to determine readabilty formulas is simple. Passages are chosen and assigned difficulty levels. Then the language within the passages is quantitatively analyzed in terms of chosen linguistic features. The difficulty level is then correlated to some ot these linguistic features and an equation is developed to represent this relationship. A correlation coefficient is calculated which represents the validity of the relationship proposed by the readability formula. Variations in the formulas occur in how passage difficulty is determined and which linguistic features are selected as variables.

To understand readability, one must understand the three components of readability research: the criterion of passage difficulty, the linguistic variables, and the correlational process that relates them. One must also be aware that the readability formulas reviewed in this paper, with the exception of the Taylor research, are predictive devices. This means that the formulas are applied to text utilizing no direct contact with readers. In contrast, the Taylor article describes a measurement device; that is, it directly involves both readers and the text. The relationship

between prediction and measurement can be seen through the discussion of the three components of readability research.

The first component consists of linguistic variables which may be extremely simple or so complecated that only a computor can handle them. Linguistic variables are characteristics of the language in the unit of text being analyzed. The linguistic variables in readability formulas are syntactic complexity and vocabulary load.

Syntactic complexity refers to the difficulty of the structure of language, which is generally analyzed on the sentence level. Research seems to indicate that complexity is related to sentence length; the longer the sentence, the more difficult. Thsu, the most common measure of syntactic complexity is simply the mean number of words per sentence. However, some research includes other systac variables such as mean number of prepositional phrases per sentence, mean words per Mean Punctuation Unit,* mean length of independent clauses, and so on, to assess the difficulty of sentences.

Vocabulary Load assesses the ease of recognition and comprehension of individual words within a passage. It is most commonly measured either by length of the words or by their absence on a word list. The mean length of words of determined by letter or syllable count, thus relating word length to the theory that the shorter a word, the more frequent and therefore more easily recognized and understood. The word list method involves counting words in a passage that do not appear in some word list. The word lists are based on frequency counts and reflect such criteria as "words most third graders can read" or "words most commonly found in pre-primer to second grade basal readers". An increase in non-list words means an increase

*A "punctuation unit" consists of all the words between any two of the following consecutive punctuation marks: comma, semi-colon, colon, period, exclamation mark, question mark, dash, parentheses.

in reading difficulty.

In statistical jargon, all the linguistic features are called predictor variables. Oddly enough, while modern readability formulas contain scores of different predictor variables (one of Bormuth's formulas has 22), all the variables can be classified under either syntactic complexity or vocabulary load. Generally, readability formulas contain one of each. Since the predictor variables must be quantifiable and are generally analyzed within short passages, other linguistic characteristics one might expect such as paragraph length, organization of ideas. literary style, etc., are not considered in most readability formulas. And since the analysis is predictove and therefore does not involve any individual characteristics of the potential readers, such factors as interests, emotions, experience, and dialectical variation cannot be considered.

The second component of readability research is the criterion of difficulty or the criterion variable. This is an estimate of the actual difficulty, expressed as a level, of each unit of text that is analyzed. This variable is of course much more difficult to assess than the predictor variables because somewhere, some form of measurement of actual readers must be included.

Researchers of readability can approach the assignment of level of text readaing difficulty in three ways, which I call direct, semi-direct, and indirect. The first approach directly utilizes readers; the second utilizes readers in a round-about way; the third utilizes only judgements about the readers.

The direct method involves a measurement of the text's readability with readers. This might involve creating a cloze test or use of a more traditional reading comprehension test from a passage, and administering this test to readers. The average age or grade level at which the passage could be successfully read and understood would become the criterion level

of that passage. For example, Lorge used normed scores on a standardized paragraph reading comprehension test to determine levels; Bormuth administered his passages as cloze tests to children to determine the passage grade levels.

In the semi-direct approach, the researcher uses another readability ranking system to assign levels to the material to be analyzed. For example, the Flesch Reading Ease formula rankings of passages were used by some researchers as criteria to determine their new formulas. Thus, the criterion variable is only another prediction rather than a measurement as in the direct method. Another variation of the semi-direct method is the use of passages from graded basal readers or textbooks. Again, the researcher relies on another prediction of level rather than a measurement. Again, the research beocmes distant from the consumers, however at least one might expect that today basals and texts are rather accurately labeled by use of several readability formulas, and that interests, relevance, ethnicity, etc., may also be considered.

The indirect method of determining reading level does not involve a direct or semi-direct measure of the readers, but instead a judgement about them by trained or knowledgeable persons. For example, the Spache formulas are based on this sort of criterion variable: books were assigned reading levels by theri availability and use in a sample of primary classrooms. Likewise, Elley used judges ranking of book difficulty as his criterion for labeling his research passages.

By now, it should be clear that the criterion variable component of the readability research and formulas is, or is based on, some form of measurement device, as discussed at the beginning of this section. We as consumers of readability formulas must trust that the researchers have in fact performed some sort of assessment or measurement to determine the stated difficulty of the passages they use. This assessment can be either directly with

readers, with the use of predictive devices based on direct contact with readers, or with the judgements of experts. It should be clear too that this is the source of the "muddy water" in readability research. Further discussion on this problem follows in the next section.

The third component of readability research is the statistical method of relating the predictor and cirterion variables. Readability formulas, called regression equations, usually look something like this:

$$Y = 1.21\ X_1 + .082X_2 + .659 \qquad (\text{Spache formula})$$

where Y is the criterion variable, most often a grade level, and the Xs represent the predictor variables. (Here X_1 is the average sentence length and X_2 is the percentage of words not on a word list.) The numbers represent the relative strengths of each variable, and the final constant adjusts the computation to the necessary grade level. An equation such as this is three-dimentional, so it is hard to visualize graphically, but basically it says, "As sentences get longer and the number of hard words increases, reading difficulty increases linearly."

A correlation estimates how accurate a regression equation really is. For example, if research reports a correlation or R = .818, it means that about 64% (R^2) of the time the equation will accurately determine readability. The other 36% of the time, other factors not named in the equation determine difficulty. Thus, of course, it is desireable for a readability formula to have as large an R as possible: the higher the R, the greater the validity of the equation.

The regression equation and the R value are determined through a statistical process called multiple correlation. R is called the multiple correlation coefficient. The value of the correlation coefficient is sometimes simple called the validity of the equation.

In order to get an idea of how multiple correlation works, suppose we

have a number of passages with assigned grade levels. We also have numerical values of several linguistic variables in each passage. We take the first predictor variable and correlate it to the criterion, and get a posotive correlation of r = .45. Thus, so far we know that the first variable accounts for about 20% of what makes the passages difficult. We correlate this with the second variable and get R = .75. Now, about half of what makes the passages difficult is known. Predictor variables are successively "thrown into the pot" (actually the "pot' is a computor) in various combinations and orders until a high R value is found. The equations is then based on the predictor variables that made this high R value. The trick is to choose the fewest variables that make the highest correlation. A simple two-variable equation with R = .80 might in the long run be desireable to a 20-variable equation with R = .99.

Note that readability equations are adjusted to represent the results of analysis of many passages at many different levels. Some readability formulas are for use only on certain levels of difficulty (Spache, Harris-Jacobson), while others proport to predict any level of readability.

In summary: Readability is a prediction of difficulty of prose. Readability is predicted by systematic use of readability formulas. These formulas are regression equations which represent how selected linguistic variables stack up against each other in relation to criteria of difficulty of the passsages which are used to determine the relationship. The higher the correlation between predictor and criterion variables, the greater the validity of the equation.

HOW GOOD IS THE RESEARCH ON READABILITY--WHAT ARE THE STRENGTHS AND LIMITATIONS?

As in the preceding section, the discussion of this question can be viewed through the 3 components of readability research: the linguistic variables. the criterion variables, and their relationship by the multiple correlation process. The problems found in each of these areas in the five readability formulas I reviewed are most likely generalizable to all readability research.

Computer technology and modern linguistic research have enabled researchers to study linguistic features of prose more and more precisely. This is most evident when we review the types of linguistic variables present in the pioneer studies as compared to those in contemporary research. For example, the Lorge and Spache variables are straight forward and easy to calculate. Basically, they include word list counts and sentence lengths--all easily counted by hand. Computer technology has enables researchers to add many more variables that would basically drive any researcher (or their lackeys) crazy trying to count manually: say the number of letters compared to the number of spaces in a 300 word passage! However, just about any variable incolving quantitative measures can be assessed by computer. While sophistication in this area is high, predictor variables of a qualitative nature have not yet been adequately considered in readability formulas. It was just this problem that Taylor addressed in the development of the cloze procedure. It is possible that in the future, as linguistic theory on semantic complexity is further developed, that the predictor variables of quantity and quality can be combined into more sophisticated readability measures. At present though, the best of the two worlds seems to be found in the Bormuth work where cloze procedure for the criterion variable plus computerized quantitative predictor variables are both used to form readability equations.

Predictor variables are the simple problem compared to criterion variables. As I read all of this research, I was constantly shocked at how reading levels of the passages were determined: many of the original formulas are directly

or indirectly based on the McCall-Crabbs multiple choice test passages which were normed in the 1920s. Others were based on teacher judgement, books found in classrooms, and text book grade leles. The only criterion that seemed directly based on actual children's work was where Bormuth used cloze tests and related these to average grade levels. The major problem with the Bormuth study is basically unsolveable--being the question, "What is comprehension?" Bormuth chose 35, 45, and 55% correct closure was indices of comprehension, but did not support any of these indices as a one baseline for comprehension. Lorge had the same problem in his criterion variable using 25, 50 and 75% correct multiple choice scores all as indices of comprehension. There really is no answer to his dilemna except arbitrariness: the researcher simply must state a minimum criterion level for "comprehension" and stick with it. An interesting corollary to this question is" do we want instruction reading level comprehension or independent reading level comprehension? I would suggest a midpoint between the two as a bench mark: say 50% closure on a cloze test or 70% comprehension/95% word recognition on an Informal Reading Inventory as arbitrary, yet happy mediums.

Another interesting problem along the lines of readability is that the formulas, in theri present state, seem to deal only at passage level. Researchers have made various statements on where to pull the passages out of a text. Some say use the beginning of a book, some say beginning, middle and end, some say find the hardest point in the book and analyze it. This last approach sounds practical to me, on the same logic that "the chain is only as strong as its weakest link"--so that a formula might give maximum difficulty. The question is, like the others, up for debate.

It is obvious to me that assigning logical, accurate levels to the passages being analyzed is a crucial aspect of readability research, and the hardest variable to pin down. My recommendations would be: 1) to use "live" subjects, as many as possible, to determine and validate average passage grade levels,

and 2) establish some logical and arbitrary measure of comprehension and explain its rationale carefully.

The third step in determination of readability formulas, application of multiple regression, is rather beyond me to criticize. Some recommendations I can make for future work (it was unclear in the articles if these were done or not) are: 1) correct the criterion variables for attenuation. This assumes that the criterion is indeed accurate, and will up the correlation and thus the reliability coefficient. 2) Choose predictor variables that correlate minimally with each other. This will simplify the resulting formula, plus give maximum power to each predictor variable used. I recommend that only 2 variables be chosen from the regression matrices for convenience to the consumer. 3) Computer is the only way to go!

In summary, the research on readability has it strengths and its problems. Klare's article points out that many researchers, some with creative approaches, have tackled the problem. In spite of all this research, only a few standard formulas have survived (Lorge, Spache, Fry) in common use, and only one relatively new formula (Harris-Jacobson) is in vogue. The most carefully researched formulas, those of Bormuth, apparently are too complicated to be in popular use. What are the problems? First, research of this nature has to be quite extensive in terms of time, materials, subjects, computer time and so on. The standard of excellence established by the Bormuth research is difficult to duplicate. Second, so far we are unable to escape the principle bugaboos of readability research: the quality vs quantity dilemna, the "What is comprehension?" question, the establishment of generalizable criterion levels. These problems are difficult and have <u>not</u>, in the research I have read, been adequately handled. This leads to a third weakness I see, that the research reports here are very vague as to methodology and materials used. While the articles are generally written for consumers of the formulas, they should be supplimented by information for fellow researchers who are trying to figure out just what their predecessors did.

COMMENTS ON THE COMPARABILITY OF THE READABILITY FORMULAS STUDIED IN THIS PROJECT

I used the formulas I reviewed on an original, ungraded story to compare their ease of use and accuracy of the formulas. On "Dreams Can Come True" I got the following reading levels:

Lorge	3.8
Spache	2.4
Bormuth (45%)	8.6
Harris-Jacobson (primary)	2.9

The results are rather disappointing to me because they vary so much. My evaluation of the story, plus that of the author, a children's literature specialist is that it is third or fourth grade level, so that I would pick the Lorge formula as the most accurate predictor here.

The Lorge formula is fairly easy to use, especially since a detailed handbook is available. Its biggest drawback is the word list, plus I disapprove of the ancient norming procedures used.

The Spache formula has several plusses. It is simple. It has rather specific delineations of grade levels for primary material. It makes no pretentions of being all encompassing. (It is for grades 1-3 only.) It has withstood the test of time, and come out to be fairly accurate by my standards.

The Elley formula is fascinating, but unfortunately, because of the word list, it requires, is impractical. I wanted to try out this formula because it has a subjective, qualitative aspect to it, but was unable to locate a suitable wordlist.

Most disappointing in terms of practical application were the Bormuth formulas. In the first place, he did not designate any one of the 24 formulas generated as being his choice, or "useful". So I picked what looked like the best manual formula, which turned out to be boring and tedious to use, and resulted in an improbable grade level, even after several recalculations. As a result, I have concluded that the Bormuth work is a model of excellent

scholarship and research, but so far of little practical value to the classroom teacher or children's librarian.

The Harris-Jacobson formulas were comparable to the Spache and Lorge. While having higher reliability that the earlier formulas, it was in my opinion, less valid. It may have an advantage though, in its more recent word list.

I think it is also interesting to look at how these formulas compare with each other in terms of relative weightings of predictor variables.

Lorge		.10V + .06 SL + .10PP + 1.99
Spache		.082V + 1.21 SL + .659
Bormuth		1.85 V + 1.31 MPU - .0001MPU^2 + 1.85
Harris-J.	P	.094V + .168SL + .502
	I	.140V + .153SL + .560

Where: V = Vocabulary depth
SL = Sentence length
MPU = Minimal Punctuation Unit

One generalization that can be made is that vocabulary is a less salient factor in formulas for lower grade materials. It is interesting that the sentence length weightings in the Harris-Jacobson formulas are so small as compared to the Spache and Bormuth formulas. Also, I can't understand what the practical use of the squared factor in Bormuth's equation is, as the constant is so small. I undertand that it makes the regression line curved, but with such a small curve, there is no practical significance.

If I were asked to recommend a readability formula for teachers to use, I would opt for the Fry graph. In truth, none of the research I studied convinced me of the superiority or accuracy of these more sophisticated formulas, so I would go for the "easiest" method of prediction. I would also ask educators to trust their own judegment. Also, I prefer not to use the wordlist methods, which are unbearable tedious, and no doubt inconsistent among readers of different backgrounds.

IMPLICATIONS: THE PRACTICAL USES OF READABILITY FORMULAS

One caution that turns up over and over again regarding readability formulas is that they are predictors, not measurements, of readability (Standal, 1978; Klare, 1974-75). People who use the formulas should keep in mind that their results are estimates, and likely to contain a good deal of error. For example--note the differences I found in the reading level of "Dreams Can Come True". I analyzed the same passage, according to the specific directions given in each formula, and came up with variability ranging over six grade levels. If we disregard the one "way off" formula (Bormuth), we have a range of grade levels 1.9 - 3.8, meaning the story is probably at second or third grade level. I believe that the formula user has to be very general about reporting or using this grade level estimate, and not get involved with precision in grade level decimal gradations or perfect adherence to the rules". The whole point I make is that the formulas provide an estimate that should be considered along with many other factors, and not regarded as "right".

A second point is--why bother with readability formulas at all? If they are so inaccurate, variable and unweildy, and if teachers and librarians can do the same job just by looking at the material, if interests and experience of the materials and the readers is not considered, really, Why bother? This is why I would bother: 1) Readability formulas do make for some kind of consistency in production and compilation of texts and children's books. They provide a theoretical baseline for naming levels on materials. With this baseline consistency, the consumers (educators) are able to make somewhat objective comparisons between texts and trade books. To summarize in broader terms, estimates of readability give general guidelines with an objective background to assignment of grade level to reading materials in the English speaking educational world. 2) Readability formulas give guidelines to noneducators: people like publishers and writers, parents, Sunday school teachers and Scout leaders. In addition, there has been some interest in readability for technical writing:

as in tax instructions, driver's manuals, information pamphlets and so on. Consideration of readability helps not only the consumers of printed materials, but also the creators.

A third implication is that it appears to me that at present, readability research has come to a standstill. For the past ten uears, after the Bormuth work, nothing really new has been done. A student of Botel, Granowsky, (cited in Klare, 1974-75) apparently tried to incorporate modern syntactic theory into a formual, but nothing was ever published. The other "new" formulas contine to be rehashes of the old ones. I see the reason for this stagnation as the current lack of availability of quantifiable semantic linguistic theory, which is the direction linguists are headed today.

The field of readability research in other languages is wide open. Some preliminary work has been done in Spanish, Chinese, Dutch, Hindi, German and so on. The possibilities of sophisticated research in this area re exciting, especially for languages of countries that are in the process of educational and literacy development. In the case of the Spanish language, three projects are readability are available (Spaulding, Patterson, and Bormuth's student). None however are in practical application. The possibility of standardizing a formula for use by educators of both first and second language Spanish readers in the United States (where the Spanish speaking people will be the largest ethnic minority group by 1985, and where Spanish is the second language) is exciting.

REFERENCES

Bormuth, John R. "Readability: A New Approach" Reading Research Quarterly, 1:3, Spring, 1966, 99-132.

Bormuth, John R. Development of Readability Analyses Final Report, Project #7-0052, Contract # OEC-3-070052-0326, USOE Bureau of Research, HEW. March, 1969, ERIC Document # ED 029-166.

Britton, Gwyneth and Margaret Lumpkin Readability Computerized Analysis: Keys to Reading Corvallis, Oregon: Britton & Associates, 1977.

Elley, Warwick B. "The Assessment of Readability by Noun Frequency Counts" Reading Research Quarterly, 4: Spring, 1969, 411-427.

Harris, Albert J. and Edward R. Sipay How to Increase Reading Ability, 6th ed. New York: David McKay, Inc. 1975, 659-675.

Klare, George A., "Assessing Readability" Reading Research Quarterly Fall, 1974, 10:1, 62-102.

Lorge, Irving R. "Predicting Reading Difficulty of Selections for Children" The Elementary English Review October, 1939, 16, 229-233.

Lorge, Irving R. "The Lorge and Flesch Readability Formulae: a Correction" School and Society February 21, 1948, 67, 141-142.

Patterson, Frank Como escribir para ser entendido. El Paso, Texas: Casa Bautista de Publicaciones, 1972.

Spache, George D. "A New Readability Formula for Primary Grade Reading Materials" Elementary School Journal, March, 1953, 53, 410-413.

Spache, George D. Good Reading for Poor Readers Champaign, Illinois: Garrard, 1974.

Spaulding, Seth "A Spanish Readability Formula" Modern Language Journal, 40, December, 1956, 433-441.

Standal, Timothy C. "Readability Formulas: What's Out, What's In?" Reading Teacher, 31:6, March, 1978, 642-646.

Taylor, Wilson L. "Cloze Procedure: A New Tool for Measuring Readability" Journalism Quarterly, 30, Fall, 1953, 415-433.